wwPDB X-ray Structure Validation Summary Report (i)

May 29, 2024 – 11:25 AM EDT

PDB ID	:	1HUC
Title	:	THE REFINED 2.15 ANGSTROMS X-RAY CRYSTAL STRUCTURE OF HUMAN LIVER CATHEPSIN B: THE STRUCTURAL BASIS FOR ITS SPECIFICITY
Authors	:	Musil, D.; Bode, W.
Deposited on	:	1993-04-21
Resolution	:	2.10 Å(reported)

This is a wwPDB X-ray Structure Validation Summary Report for a publicly released PDB entry.

We welcome your comments at *validation@mail.wwpdb.org*
A user guide is available at
https://www.wwpdb.org/validation/2017/XrayValidationReportHelp
with specific help available everywhere you see the (i) symbol.

The types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

The following versions of software and data (see references (i)) were used in the production of this report:

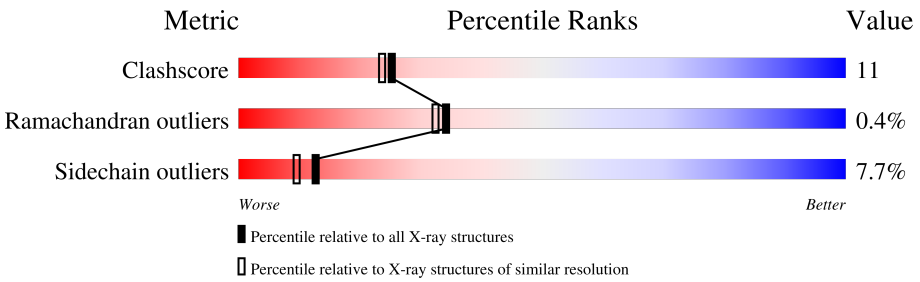
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	NOT EXECUTED
EDS	:	NOT EXECUTED
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36.2

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 2.10 Å.

Percentile scores (ranging between 0-100) for global validation metrics of the entry are shown in the following graphic. The table shows the number of entries on which the scores are based.



Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
Clashscore	141614	5710 (2.10-2.10)
Ramachandran outliers	138981	5647 (2.10-2.10)
Sidechain outliers	138945	5648 (2.10-2.10)

The table below summarises the geometric issues observed across the polymeric chains and their fit to the electron density. The red, orange, yellow and green segments of the lower bar indicate the fraction of residues that contain outliers for >=3, 2, 1 and 0 types of geometric quality criteria respectively. A grey segment represents the fraction of residues that are not modelled. The numeric value for each fraction is indicated below the corresponding segment, with a dot representing fractions <=5%

Note EDS was not executed.

Mol	Chain	Length	Quality of chain
1	A	47	72% 21% 6%
1	C	47	66% 28% 6%
2	B	205	62% 31% 6%
2	D	205	60% 35% .

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 4159 atoms, of which 0 are hydrogens and 0 are deuteriums.

In the tables below, the ZeroOcc column contains the number of atoms modelled with zero occupancy, the AltConf column contains the number of residues with at least one atom in alternate conformation and the Trace column contains the number of residues modelled with at most 2 atoms.

- Molecule 1 is a protein called CATHEPSIN B.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	47	364	225	65	70	4	10	0	0
1	C	47	364	225	65	70	4	12	0	0

- Molecule 2 is a protein called CATHEPSIN B.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	B	205	1574	988	265	307	14	38	0	0
2	D	205	1574	988	265	307	14	45	0	0

- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
3	A	32	32	32	0	0
3	B	121	121	121	0	0
3	C	27	27	27	0	0
3	D	103	103	103	0	0

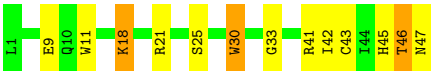
3 Residue-property plots (i)

These plots are drawn for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry. Residues are color-coded according to the number of geometric quality criteria for which they contain at least one outlier: green = 0, yellow = 1, orange = 2 and red = 3 or more. Stretches of 2 or more consecutive residues without any outlier are shown as a green connector. Residues present in the sample, but not in the model, are shown in grey.

Note EDS was not executed.

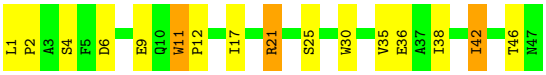
• Molecule 1: CATHEPSIN B

Chain A: 72% 21% 6%



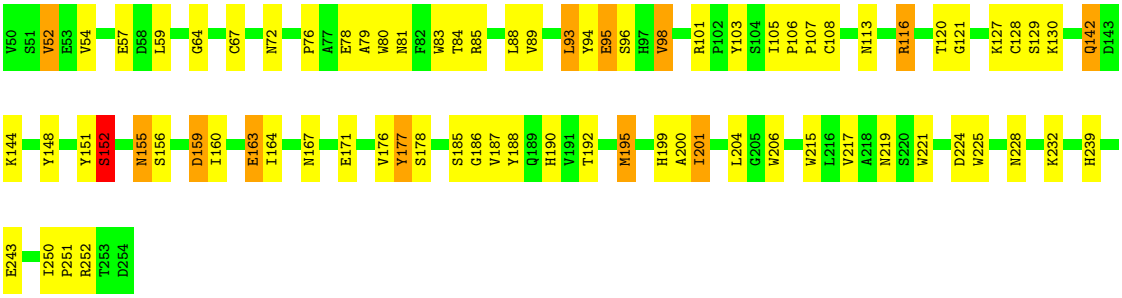
• Molecule 1: CATHEPSIN B

Chain C: 66% 28% 6%



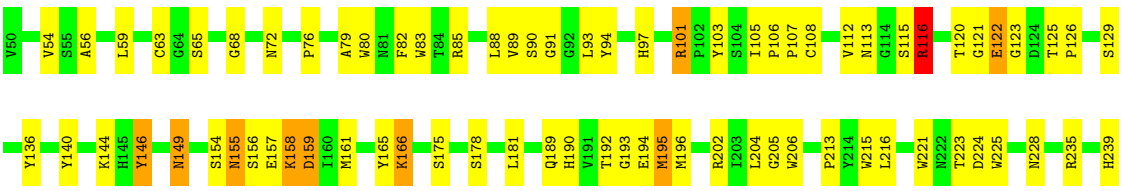
• Molecule 2: CATHEPSIN B

Chain B: 62% 31% 6%



• Molecule 2: CATHEPSIN B

Chain D: 60% 35%



C240
G241
I242
E243
S244
E245
V246
V247
A248
D254

4 Data and refinement statistics (i)

Xtriage (Phenix) and EDS were not executed - this section is therefore incomplete.

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	86.23Å 34.16Å 85.56Å 90.00° 102.90° 90.00°	Depositor
Resolution (Å)	(Not available) – 2.10	Depositor
% Data completeness (in resolution range)	(Not available) ((Not available)-2.10)	Depositor
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
Refinement program	EREF, X-PLOR	Depositor
R, R_{free}	0.164 , (Not available)	Depositor
Estimated twinning fraction	No twinning to report.	Xtriage
Total number of atoms	4159	wwPDB-VP
Average B, all atoms ($Å^2$)	16.0	wwPDB-VP

5 Model quality (i)

5.1 Standard geometry (i)

The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 5$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	$\#\|Z\| >5$	RMSZ	$\#\|Z\| >5$
1	A	1.08	2/373 (0.5%)	1.54	2/505 (0.4%)
1	C	1.12	2/373 (0.5%)	1.64	3/505 (0.6%)
2	B	1.17	6/1622 (0.4%)	1.61	15/2203 (0.7%)
2	D	1.12	6/1622 (0.4%)	1.51	12/2203 (0.5%)
All	All	1.14	16/3990 (0.4%)	1.57	32/5416 (0.6%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand.A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
1	A	0	1
1	C	0	1
2	B	0	10
2	D	0	15
All	All	0	27

The worst 5 of 16 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	D	80	TRP	NE1-CE2	-8.41	1.26	1.37
2	D	221	TRP	NE1-CE2	-8.25	1.26	1.37
2	B	206	TRP	NE1-CE2	-8.02	1.27	1.37
1	C	11	TRP	NE1-CE2	-7.66	1.27	1.37
2	B	83	TRP	NE1-CE2	-7.54	1.27	1.37

The worst 5 of 32 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	C	21	ARG	NE-CZ-NH2	-14.80	112.90	120.30
2	B	116	ARG	NE-CZ-NH2	-14.19	113.21	120.30
1	C	21	ARG	NE-CZ-NH1	12.23	126.42	120.30

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	B	116	ARG	NE-CZ-NH1	9.84	125.22	120.30
2	B	101	ARG	NE-CZ-NH1	-8.73	115.94	120.30

There are no chirality outliers.

5 of 27 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	21	ARG	Sidechain
2	B	54	VAL	Mainchain
2	B	57	GLU	Mainchain
2	B	64	GLY	Mainchain
2	B	88	LEU	Mainchain

5.2 Too-close contacts (i)

In the following table, the Non-H and H(model) columns list the number of non-hydrogen atoms and hydrogen atoms in the chain respectively. The H(added) column lists the number of hydrogen atoms added and optimized by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, whereas Symm-Clashes lists symmetry-related clashes.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	364	0	340	8	0
1	C	364	0	340	8	0
2	B	1574	0	1444	35	0
2	D	1574	0	1444	33	0
3	A	32	0	0	0	0
3	B	121	0	0	8	0
3	C	27	0	0	1	0
3	D	103	0	0	2	1
All	All	4159	0	3568	77	1

The all-atom clashscore is defined as the number of clashes found per 1000 atoms (including hydrogen atoms). The all-atom clashscore for this structure is 11.

The worst 5 of 77 close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
2:D:120:THR:HG22	2:D:121:GLY:H	1.37	0.87
2:B:190:HIS:H	2:B:239:HIS:HE1	1.29	0.81

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:43:CYS:O	1:A:46:THR:HG22	1.88	0.73
2:B:96:SER:HB2	2:B:98:VAL:HG13	1.72	0.72
2:D:242:ILE:HG23	2:D:243:GLU:HG3	1.76	0.66

All (1) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
3:D:747:HOH:O	3:D:776:HOH:O[1_565]	2.17	0.03

5.3 Torsion angles (i)

5.3.1 Protein backbone (i)

In the following table, the Percentiles column shows the percent Ramachandran outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the backbone conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	A	45/47 (96%)	44 (98%)	1 (2%)	0	100	100
1	C	45/47 (96%)	45 (100%)	0	0	100	100
2	B	203/205 (99%)	193 (95%)	9 (4%)	1 (0%)	29	26
2	D	203/205 (99%)	185 (91%)	17 (8%)	1 (0%)	29	26
All	All	496/504 (98%)	467 (94%)	27 (5%)	2 (0%)	34	32

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	B	121	GLY
2	D	65	SER

5.3.2 Protein sidechains (i)

In the following table, the Percentiles column shows the percent sidechain outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar

resolution.

The Analysed column shows the number of residues for which the sidechain conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
1	A	39/39 (100%)	36 (92%)	3 (8%)	13 9
1	C	39/39 (100%)	37 (95%)	2 (5%)	24 22
2	B	169/169 (100%)	155 (92%)	14 (8%)	11 7
2	D	169/169 (100%)	156 (92%)	13 (8%)	13 9
All	All	416/416 (100%)	384 (92%)	32 (8%)	13 9

5 of 32 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	D	155	ASN
2	D	195	MET
2	B	178	SER
2	B	163	GLU
2	D	204	LEU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 7 such sidechains are listed below:

Mol	Chain	Res	Type
2	D	149	ASN
2	D	155	ASN
2	D	239	HIS
2	D	237	GLN
2	B	239	HIS

5.3.3 RNA (i)

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains (i)

There are no non-standard protein/DNA/RNA residues in this entry.

5.5 Carbohydrates (i)

There are no monosaccharides in this entry.

5.6 Ligand geometry (i)

There are no ligands in this entry.

5.7 Other polymers (i)

There are no such residues in this entry.

5.8 Polymer linkage issues (i)

There are no chain breaks in this entry.

6 Fit of model and data (i)

6.1 Protein, DNA and RNA chains (i)

EDS was not executed - this section is therefore empty.

6.2 Non-standard residues in protein, DNA, RNA chains (i)

EDS was not executed - this section is therefore empty.

6.3 Carbohydrates (i)

EDS was not executed - this section is therefore empty.

6.4 Ligands (i)

EDS was not executed - this section is therefore empty.

6.5 Other polymers (i)

EDS was not executed - this section is therefore empty.